Course: CS699 A2 – Data Mining

Assignment: Project Assignment No.1 – Proposal

Team: Mike Zhong, Jan Allemann

# Dataset

The dataset, obtained from INSERTSOURCE, contains 350 tuples with 27 attributes each. The attributes are described in the following table:

|  |  |  |
| --- | --- | --- |
| Attribute | Format | Description |
| SALE.TYPE | Categorical |  |
| SOLD.DATE | Datetime |  |
| PROPERTY.TYPE | Categorical |  |
| ADDRESS | String |  |
| CITY | String |  |
| STATE.OR.PROVINCE | String |  |
| ZIP.OR.POSTAL.CODE | Integer |  |
| PRICE | Integer |  |
| BEDS | Integer |  |
| BATHS | Integer |  |
| LOCATION | String |  |
| SQUARE.FEET | Integer |  |
| LOT.SIZE | Integer |  |
| YEAR.BUILT | Integer |  |
| DAYS.ON.MARKET | Integer |  |
| X..SQUARE.FEET | Integer |  |
| HOA.MONTH | Categorical |  |
| STATUS | Categorical |  |
| NEXT.OPEN.HOUSE.START.TIME | Datetime |  |
| NEXT.OPEN.HOUSE.END.TIME | Datetime |  |
| URL | String |  |
| SOURCE | Categorical |  |
| MLS | Integer |  |
| FAVORITE | Binary |  |
| INTERESTED | Binary |  |
| LATITUTDE | Numerical |  |
| LONGITUDE | Numerical |  |

# Classification Goal

The class attribute is the PROPERTY.TYPE, which has the following 7 categories:

Condo/Co-op

Single Family Residential

Townhouse

Multi-Family (2-4 Unit)

Vacant Land

Mobile/Manufactured Home

Multi-Family (5+ Unit)

The goal of the project is, to predict the PROPERTY.TYPE category.